

Predicting Forest Stand Properties from Satellite Images with Different Data Mining Techniques

Katerina Taškova¹, Panče Panov², Andrej Kobler³, Sašo Džeroski², Daniela Stojanova³

¹Faculty of Information Technology, European University, Skopje, Macedonia

²Department of Knowledge Technologies, Jožef Stefan Institute, Ljubljana, Slovenia.

³Slovenian Forestry Institute, Ljubljana, Slovenia

E-mails: tkejt@yahoo.com,

{Pance.Panov, Saso.Dzeroski}@ijs.si,

Andrej.Kobler@gozdis.si, stojanovad@yahoo.com

Abstract

This paper work is focused on the comparison of different data mining techniques and their performances by building predictive models of forest stand properties from satellite images. We used the WEKA data mining environment to implement our numeric prediction experiments, applying linear regression, model (regression) trees, and bagging. The best results (with regard to correlation) we obtained by bagging model trees for considered target attributes.

1 Introduction

The main idea for this study is to supply more consistent and more accurate supporting information to the forest monitoring system, in order to predict the accumulated forest biomass. The Slovenian Forestry Service operates such a monitoring system [2], which periodically provides a wide range of forestry related information using an extensive network of permanent field sample plots throughout Slovenia. This system is tested and it is proven to be reliable, but it is also very labor-intensive and costly. Furthermore, some of the forest stand attributes, such as canopy cover, can only be roughly estimated by visual observation. Other items, such as forest stand height, can be monitored only seldom due to technical difficulties of field measurements.

The work in this paper is focused on the comparison of different data mining techniques and their performances by building predictive models of forest stand properties from satellite images. These predictive models were formed on multi-temporal Landsat data, calibrated with remotely sensed data, acquired by very high resolution airborne laser scanning (ALS), also termed LIDAR (Light Detection and Ranging).

LIDAR is one of many laser remote sensing techniques [3] that is used in forestry for estimation of different parameters. Because of its immediate generation of 3D data, high spatial resolution (in the order of a few centimeters) and accuracy, ALS data is becoming popular for detailed

measurements of forest stand height and estimating other forest stand parameters [1]. By model-based extrapolation from a lidar-scanned spatial sample to a wide area coverage conventional optical remote sensing products such as those from the Landsat Enhanced Thematic Mapper Plus (Landsat ETM+) sensor can play an important role in bringing lidar into main-stream forestry practice.

Landsat Thematic Mapper (TM) is a multispectral scanning radiometer that was carried on board Landsats 4 and 5. The Landsat Enhanced Thematic Mapper (ETM) was introduced with Landsat 7. ETM data cover the visible, near-infrared, shortwave, and thermal infrared spectral bands of the electromagnetic spectrum.

The description of the data together with the main target attributes used in our prediction analyses are presented in Section 2. Overview of the data mining algorithms we used in the analyses for generating predictive models is given in Section 3. In Section 4 we give the experimental setup. In Section 5 we present and discuss the experimental results and in Section 6 we present the conclusions and give pointers to future work.

2 Data description

The study area encompassed 72226 hectares of the Kras region in western Slovenia. It is covered by a highly fragmented landscape of forests, shrubs and pastures. The forests contain mostly oak (*Quercus pubescens*) and pine (*Pinus nigra*) of various ages and stand compositions. Multi spectral Landsat ETM+ data were acquired on August 3rd 2001, May 18th 2002, November 10th 2002, and March 18th 2003, thus capturing the main phenological stages of forest vegetation in the area. The Landsat imagery was first geometrically corrected by orthorectification. Each of the 4 Landsat images was then segmented at two levels of spatial detail. The average image segment sizes were 4 ha for the fine segmentation and 20 ha for the coarse segmentation. Based on the data within each image segment 4 statistics (minimum reflectance, maximum reflectance, average reflectance, standard deviation of reflectance) were computed for each date, for each segmentation level, and

for each of the Landsat image channels (2, 3, 4, 5, 7) and this way 160 explanatory variables were derived for forest modeling. As the borders of individual segments were not identical between dates and segmentation levels, all 160 variables values were attributed back to individual image pixels.

The 11 target variables describing the forest were computed at the level of 25 m by 25 m squares from the LIDAR data set, corresponding to Landsat pixels. For the purposes of our analyses we used only two target variables - canopy cover and forest stand height. Therefore, we will mention and describe them briefly in the rest of this section.

The forest stand height (FSH) for each square (or Landsat pixel) was computed by averaging the heights of the LIDAR-based normalized digital surface model (nDSM) within the 25 m square. A nDSM is a high resolution raster map showing the relative height of vegetation above the bare ground.

The canopy cover (CC) within this study is defined as the percentage of bare ground within a 25 m square (or a Landsat pixel), covered by a vertical projection of the overlying vegetation, higher than 1 m.

3 Data mining methodology

According to our motivation for this study, our aim was to see how other/different data mining techniques would apply to the given dataset in order to improve the results gained from previous analyses [7] with the same data.

This time the analyses were made only in the WEKA environment, concentrating on the learning schemes and their performances that would best fit to our problem. Because of the numeric nature of the data, we used linear regression, model and regression trees, and bagging.

3.1 Linear Regression

When we are dealing with numeric prediction problem, as is the case in this study, linear regression is the most natural technique to consider. The idea is simple: to represent the class (target attribute) x as linear function of the attributes a_1, \dots, a_k with predetermined weights w_0, \dots, w_k

$$\begin{aligned} x &= w_0 + w_1 a_1 + w_2 a_2 + \dots + w_k a_k = \\ &= w_0 a_0 + w_1 a_1 + w_2 a_2 + \dots + w_k a_k, \end{aligned}$$

if we assumed $a_0 = 1$.

The method of linear regression is basically the appropriate choice of the weight coefficients, w_j $j=0, \dots, k$, from the given training data. Corresponding ones should minimize the sum

$$\sum_{i=1}^n (x^{(i)} - y^{(i)})^2, \quad y^{(i)} = \sum_{j=0}^k w_j a_j^{(i)}$$

of the squares of the differences between the actual $x^{(i)}$ and the predicted $y^{(i)}$ values of the class of the i -th instance over whole training set. [6]

Linear regression is simple and basic numeric predictive scheme, which fits very good in many statistical analyses for decades. Because of the linear representation of the class value, is obviously that even this model is very simple, it has the disadvantage of suffering from linearity. Not every dataset can follow linear dependency. In that case the best-fitting line that can be found in order of least mean-squared difference will represent the class.

3.2 Regression and model trees

Regression and model trees [4] both derive from the basic divide-and-conquer method for decision tree construction. Once the basic tree has been build, consideration is given to backward tree pruning from each leaf as with ordinary decision trees [6]. Exception from the ordinary decision trees is the value stored at each terminal node (leaf). If the leaf is represented with single predicted numeric value-average value of the class value of the instances that reach the leaf, we have described regression trees. If we combine regression trees with the simple linear regression method (subsection 3.1), we get a more sophisticated representation of regression trees called model trees. These trees include linear regression model for class value prediction of the instances that reach the particular leaf. Model trees are smaller, more comprehensible and perform better prediction accuracy compared with regression trees.

3.3 Bagging

Combining multiple models outputs, in order to make decisions more reliable, is very logical approach in data mining methodology that has been used very widely. Several machine learning techniques, among them bagging, apply this approach by learning an ensemble of models and using them in combination. Despite the disadvantage of being difficult to analyze, often their improvements in the predictive performances compared to single model ones, is one of the key-reasons for their successful employment both in numeric prediction and classification tasks.

Bagging (stands for "bootstrap aggregating") is complex learning method that combines two concepts: bootstrapping [9] and aggregating presented by Breiman [5]. Bootstrapping is a sampling procedure based on random selection with replacement. When applied to the original training set $X = (X_1, X_2, \dots, X_n)$ with n instances, it alters it by deleting some instances and replicating others in order to create a new one $X^b = (X_1^b, X_2^b, \dots, X_n^b)$ with the same size. Bagging then learns predictive models over these artificially derived training sets and uses them to generate an aggregated predictor. Aggregating actually means combining of classifiers. The aggregation averages

over the models outcome when predicting a numeric target attribute and does plurality vote when predicting a class [5].

The bootstrap procedure is obviously a significant factor that contributes in bagging performances, because in this way can be avoided the possible outliers (instances that misrepresent the real data distribution) in the original training set. Therefore, bagging is helpful for building better classifiers on training sets with outliers. In bagging, bootstrapping and aggregating techniques are implemented in the following way:

1. Repeat for $b = 1, \dots, B$
 - a) Take a bootstrap replicate X^b of the training dataset X .
 - b) Construct a base classifier $C^b(x)$ on X^b .
2. Combine base classifiers $C^b(x)$, $b = 1, \dots, B$, by the simple majority vote (or by averaging their outcomes) to a final prediction.

Considering Breiman's analyses, bagging could give substantial accuracy improvements when applied on unstable prediction methods as classification and regression trees are.

4 Experimental setup

As described before, we learn to make predictions about the forest stand properties by using Landsat images. The prediction task consists of building predictive models by using data mining algorithms and validating the models by using standard validation techniques.

The WEKA [6] workbench includes a wide collection of machine learning algorithms and provides an environment that enables WEKA users to test and compare different learning techniques over their performances. Due to our main idea, we used the WEKA data mining environment to implement our numeric prediction experiments for comparison of the learning algorithms described in previous section.

All experiments we made were on the dataset described in detail in Section 2. The dataset consisted of 160 descriptive attributes and 11 target attributes of which two (forest stand height (FSH) and canopy cover (CC)) were used for the prediction task. All descriptive and target attributes were numeric. The dataset has 64000 examples of which 60607 describe the vegetation outside a settlement and were used for building the predictive models.

We used WEKA's Linear Regression algorithm, as a baseline method, to build linear regression model for CC and FSH with default parameters. We compared the baseline method with WEKA's implementation of the M5 algorithm for learning model (regression) trees. In order to prevent trees from over fitting the data, we employed pruning during tree construction. We altered different values (2^n , $n=2\dots 10$) of the pruning parameter (minimal number of instances per

leaf) in order so see how this affects the correlation and the size of the models.

The last experiment involved building predictive models with WEKA's implementation of the bagging algorithm. As base-level algorithms we used M5 regression and model trees.

Because of the large dimensionality of the analyzed dataset (160 descriptive attributes), considered learning algorithms were very time consuming. Hence we considered attribute selection by employing the instance-based attribute evaluator Relief to rank the descriptive attributes and lower the dimensionality of the data.

The performance of all models was validated using 10-fold cross-validation.

5 Results and discussion

In this section, we present the results we obtained from the experiments. As it was explained previously, our main target attributes, canopy cover (CC) and forest stand height (FSH), are numeric; therefore we used WEKA's implementation of linear regression, model and regression trees and bagging algorithms to obtain the predictive models. In Table 1 we present the correlation measure for the obtained predictive models for both target attributes.

Data Mining Method	Correlation			
	Canopy Cover		Forest Stand Height	
	All attributes (160)	Attribute selection (50)	All attributes (160)	Attribute selection (50)
Linear regression	0.836	0.819	0.814	0.7699
Regression tree (default)	0.858	0.852	0.877	0.864
Model tree (default)	0.863	0.857	0.886	0.875
Bagging of regression trees	0.876	0.869	0.893	0.884
Bagging of model trees	0.882	0.871	0.902	0.891

Table 1. Correlation measure for the obtained prediction models of Canopy Cover and Forest Stand Height

The first column is the WEKA's models we build. The next four represent the correlation coefficients obtained for canopy cover and forest stand height prediction models respectively. From the results we can conclude that linear regression model with default parameters has the lowest correlation compared to M5 model and regression trees for both attributes. The best results (with regard to correlation) we obtained by bagging model trees for both target attributes. Obviously, application of attribute selection over original dataset leads to slightly smaller correlation values compared to ones obtained from the same learning methods

applied to the original dataset. This could be very important approach when analyzing large dataset in order to speed up used learning algorithms without substantial affection on the model accuracy.

The results in Table 2 present the dependency of the correlation of the M5 model trees and the number of leaves (number of rules) by applying different degrees of pruning. This is also presented graphically in Figure 1. From the results we can see that if we increase the pruning parameter (minimal number of instances in a leaf) the correlation is constant to some degree of pruning and then starts to decrease. By increasing the pruning parameter the size of the trees is also decreasing. If we want our models to be smaller we have to compensate with the accuracy of the model.

Minimal number of instances per leaf	Canopy Cover		Forest Stand Height	
	Correlation	Num. of Leaves	Correlation	Num. of Leaves
4	0.863	1036	0.886	1220
8	0.863	1263	0.886	1199
16	0.863	1221	0.886	1110
32	0.861	1025	0.882	974
64	0.858	770	0.878	753
128	0.853	545	0.871	493
256	0.847	314	0.860	256
512	0.841	92	0.849	121
1024	0.836	44	0.836	49

Table 2. Correlation and number of leaves in the M5 model trees obtained by different degrees of pruning

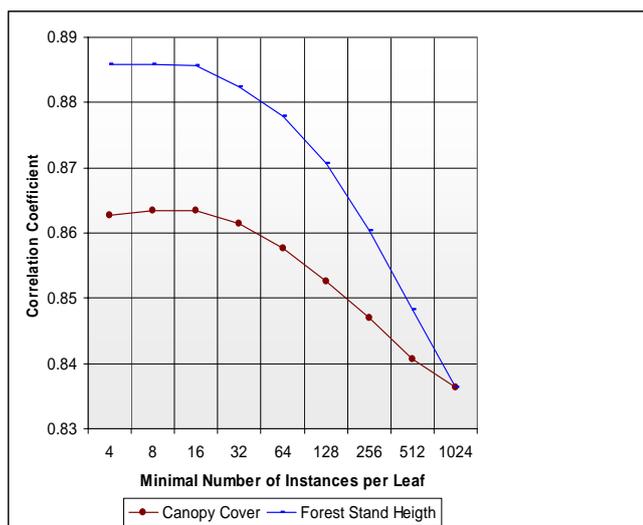


Figure 1. Graphical representation of the dependency of the correlation of M5 model trees from the degree of pruning

6 Conclusions and Future Work

In this paper we compared a number of predictive methods: linear regression, regression and model trees and bagging of regression and model trees. To lower the dimensionality of the data and to try to improve the results obtained in the previous study [7] we applied feature selection algorithms. From the results of the performed experiments we can conclude that linear regression gives the worst results and that bagging of model trees improves the results obtained in [7]. Also performed feature selection of the data doesn't help to improve the performances of the models.

References

- [1] Hyypä, H. (et al.): Algorithms and Methods of Airborne Laser-Scanning for Forest Measurements. International Archives of Photogrammetry and Remote Sensing, Vol XXXVI, 8/W2, Freiburg, Germany, (2004)
- [2] SFS Slovenian Forestry Service: Slovenian forest and forestry. Zavod za gozdove RS, 24 pp (1998)
- [3] Raymond, M. (et al.): Measures. Laser remote sensing: fundamentals and applications. Malabar, Fla., Krieger Pub. Co. 510 p. G70.6.M4 (1992)
- [4] Quinlan, J. R.: Learning with continuous classes. In Proceedings of the 5th Australian Joint Conference on Artificial Intelligence, pages 343–348. World Scientific, Singapore, (1992).
- [5] Breiman, L.: Bagging Predictors. Machine Learning Vol. 26(2) 123–140, (1996)
- [6] Witten, I. (et al.): Data Mining: Practical machine learning tools and techniques with Java implementations. Second Edition, Morgan Kaufman, (2005)
- [7] Džeroski, S. (et al.): Using decision trees to predict forest stand height and canopy cover from LANSAT and LIDAR data. In: Managing environmental knowledge: EnviroInfo 2006: proceedings of the 20th International Conference on Informatics for Environmental Protection, Aachen: Shaker Verlag, pg. 125-133, (2006)
- [8] Efron, B. (et al.): An Introduction to the Bootstrap. Chapman and Hall, New York, (1993)